SOLARDATA: UN DATASET EXHAUSTIVO DE PLANTAS SOLARES. CASO DE USO: LOCALIZACIÓN ÓPTIMA DE PLANTAS PV

A PREPRINT

Anibal Mantilla-Guerra
FIGEMPA
Central University of Ecuador
Quito, Ecuador
armantilla@uce.edu.ec

Christian Mejia-Escobar
FIGEMPA
Central University of Ecuador
Quito, Ecuador
cimejia@uce.edu.ec

Jorge Azorin-Lopez
Department of Computer Technology
University of Alicante
Alicante, Spain
jazorin@ua.es

Jose Garcia-Rodriguez
Department of Computer Technology
University of Alicante
Alicante, Spain
jgarcia@dtic.ua.es

Byron Tarco
Department of Geology
Central University of Ecuador
Quito, Ecuador
bftarco@uce.edu.ec

Karen Santamaria
Department of Geology
Central University of Ecuador
Quito, Ecuador
knsantamaria@uce.edu.ec

January 15, 2026

ABSTRACT

El uso de la energía limpia es una tendencia global, y la generación eléctrica mediante centrales solares fotovoltaicas es un pilar fundamental de esta transición energética, experimentando un crecimiento constante. Para optimizar su aprovechamiento y múltiples aplicaciones, es esencial disponer de datos solares detallados; sin embargo, los conjuntos de datos actuales son limitados y no logran combinar los diversos factores clave que influyen en su desarrollo. Este estudio presenta una metodología replicable para la recopilación, generación e integración de datos coherentes y fiables sobre las plantas solares fotovoltaicas a nivel mundial. Utilizando datos coherentes, de alta resolución y fiables, así como una metodología replicable, este estudio presenta: Nuestro trabajo aborda esta limitación al introducir un dataset integral sobre plantas solares, compuesto por 25 variables de naturaleza cualitativa y cuantitativa, que abarcan las características naturales, técnicas, sociales y económicas de instalaciones en todos los países del mundo. Este nuevo recurso, que es novedoso, cuidadosamente elaborado y públicamente disponible, constituye un aporte primordial para estudios e investigaciones en el campo energético. Finalmente, demostramos la utilidad de este conjunto de datos para desarrollar una aplicación de inteligencia artificial destinada a la localización óptima de plantas solares mediante deep learning, para ser usada por científicos o público en general interesado en energía solar.

Keywords Solar PV power · Dataset · Deep Learning · Energy

1 Introducción

El uso de la energía se ha convertido en una preocupación crucial a nivel mundial debido al incremento de la demanda energética, la necesidad de mitigar el cambio climático y el calentamiento global, así como a los graves impactos ambientales derivados de la generación convencional y a la urgencia de reducir la dependencia de los combustibles fósiles. En este contexto, la búsqueda de alternativas más limpias y sostenibles ha impulsado el desarrollo de fuentes de electricidad verde, destacándose la energía solar como un pilar fundamental de la transición energética global [1][2].

La radiación solar es esencialmente un recurso gratuito disponible en cualquier lugar de la Tierra, en mayor o menor medida. Las plantas fotovoltaicas (PV) convierten la radiación solar en electricidad y son el tipo más común de instalación de energía solar [3]. Su despliegue estratégico permite diversificar la matriz energética, mejorar la seguridad

del suministro y reducir el impacto ambiental. Estas infraestructuras se han consolidado como una solución madura y viable gracias a los avances tecnológicos y a la constante reducción de sus costos de fabricación [4].

La tecnología fotovoltaica se convierte en una oportunidad para que los países y las comunidades transformen o desarrollen su infraestructura energética. Por su creciente y acelerada expansión en todo el mundo, resulta esencial la creación de un inventario completo y actualizado de plantas de energía solar. Este registro no solo permite realizar un seguimiento del progreso y despliegue de la energía solar, sino que es un paso crítico para resolver un problema clave dentro de este sector: elegir los lugares adecuados para el emplazamiento de proyectos de energía solar. Esta tarea es fundamental para respaldar el crecimiento de las energías renovables, la planificación estratégica, lograr el desarrollo energético sostenible y fomentar la toma de decisiones informadas en materia de políticas y estrategias de inversión.

El problema de localización óptima de plantas solares busca encontrar los puntos geográficos donde una planta produciría más energía, al menor costo y con la mayor viabilidad técnica y ambiental. Es una actividad inherentemente compleja y constituye un desafío multidisciplinario que requiere considerar factores técnicos, operativos, económicos, ambientales y sociales. Un inventario de plantas solares proporcionaría los datos reales que describen cómo se comportan las plantas existentes, convirtiéndose en una base empírica para desarrollar y validar cualquier tipo de método de localización.

Aunque existen datos relacionados con la energía solar, la creación de un inventario preciso de las plantas de energía solar enfrenta varios retos. La recopilación eficaz de datos depende de la integración de diversos conjuntos de datos procedentes de diversas fuentes, como organismos gubernamentales y organizaciones privadas. Los conjuntos de datos actuales son limitados, heterogéneos o inexistentes en diversas regiones. Los datasets disponibles se encuentran aislados y fragmentados, algunos proporcionan información parcial, pero ninguno integra de forma estandarizada los atributos de todos los factores influyentes en el aprovechamiento de la energía solar. Esta carencia de datos integrados, completos y globales constituye un vacío científico evidente que motivó la creación de un conjunto de datos amplio y de calidad que reúna los atributos esenciales para el análisis de emplazamiento fotovoltaico.

El análisis comparativo de la literatura relacionada permitió identificar variables de diversas categorías, las cuales son indispensables para construir un modelo robusto de selección de sitios. Un dataset de calidad exige incorporar información adicional sobre la irradiación solar y el clima, la topografía e idoneidad del terreno, la proximidad a la infraestructura vial y eléctrica, la sensibilidad ambiental y los impactos socioeconómicos [5]. Se planteó una metodología híbrida basada en el procesamiento geoespacial de Modelos Digitales de Elevación (DEM) oficiales procesados mediante algoritmos reproducibles, complementada con la adquisición de datos climáticos y datos de centrales de generación solar fotovoltaica. Se obtuvieron desde fuentes de prestigio un total de 27 variables y miles de registros de todos los países del mundo. En esta tarea, los Sistemas de Información Geográfica (SIG) desempeñan un papel fundamental, ya que permiten integrar diversas capas de información [6]. Esta combinación permitió integrar factores geográficos, topográficos, climáticos y técnicos en un único dataset.

En este estudio, la utilidad práctica de este dataset es demostrada mediante un modelo predictivo de deep learning basado en redes neuronales para la estimación del potencial PV y la generación automática de mapas de idoneidad, lo cual se convierte en un sistema de recomendación para emplazamiento óptimo.

En conclusión, este trabajo aporta 3 contribuciones principales: i) un conjunto de datos de calidad, fiable y libre acceso sobre plantas solares PV; ii) un flujo de trabajo reproducible para la recopilación, organización y depuración de grandes volúmenes de datos, adaptable a otros campos. Aplica procedimientos de limpieza, depuración y estandarización y verifica consistencia interna entre variables; y iii) una aplicación basada en deep learning para la localización óptima de plantas solares fotovoltaicas

De tal forma, nuestro propósito es proporcionar herramientas basadas en datos, modelos y procedimientos útiles para la comunidad científica, tanto principiantes como expertos, para avanzar en investigaciones sobre la selección óptima de sitios, el diseño de plantas solares y la predicción de rendimiento. También puede servir como herramienta para atraer inversiones del sector público y/o privado, ya que apoya el proceso de toma de decisiones informadas basadas en información precisa y actualizada relativa a los proyectos de infraestructura energética.

El resto del documento se estructura de la siguiente manera: la Sección 2 explora los trabajos relacionados sobre datasets y modelos acerca de energía solar. La Sección 3 ofrece una descripción detallada de nuestro conjunto de datos y del proceso diseñado para su creación. La Sección 4 realiza un análisis estadístico de los atributos cualitativos y cuantitativos para obtener conocimiento. Luego, presentamos un caso de uso del dataset creado en la Sección 5. La Sección 6 detalla nuestras conclusiones y el trabajo futuro.

2 Trabajos relacionados

En esta sección se presentan los trabajos más estrechamente relacionados y se analizan sus principales aportes, metodologías y limitaciones. Se organiza en dos partes complementarias. En primer lugar, se revisan los principales datasets existentes sobre plantas solares, considerando su alcance, nivel de detalle y disponibilidad pública. En segundo lugar, se analizan los modelos y metodologías utilizados para la localización óptima de instalaciones fotovoltaicas, incluyendo enfoques basados en SIG, métodos multicriterio y técnicas de aprendizaje automático.

2.1 Datasets existentes

Nuestra revisión de la literatura sobre las bases de datos más comúnmente utilizadas en estudios del clima y proyectos solares se resume en la Tabla 1.

Table 1: Resumen de bases de datos de radiación solar.

#	Base de datos	Acceso	Variables principales	Período	Aplicaciones
1	NASA POWER	Gratis	GHI, DNI, DHI, temp., viento, nubes	1983–actual	Preliminares, IA
2	PVGIS	Gratis	GHI, DNI, difusa, albedo, temp.	2005–actual	Diseño PV, simulaciones
3	Global Solar Atlas	Gratis/Pago	GHI, DNI, difusa, temp., productible	>20 años	Plantas solares
4	SolarGIS	Pago	GHI, DNI, clima, suciedad, degradación	>20 años	Ingeniería, gemelos digitales
5	Meteonorm	Pago (demo)	GHI, DNI, clima completo	Sintético	Rellenar huecos, simulación
6	ERA5 (Copernicus)	Gratis	GHI, nubes, viento, temp.	1979–actual	Modelos predictivos, IA
7	MERRA-2 (NASA)	Gratis	GHI, nubes, aerosoles	1980–actual	IA y clima
8	CERES (NASA)	Gratis	Radiación entrante/saliente, nubes	2000–actual	Balance energético
9	HelioClim (HC3/HC4)	Gratis/Pago	GHI, DNI	2004–actual	Series homogéneas PV
10	BSRN	Gratis	Radiación medida en superficie	1992–actual	Alta precisión
11	SURFRAD (NOAA)	Gratis	Radiación y clima	1995–actual	Medición precisa

Es necesario destacar que la radiación solar se clasifica principalmente en tres componentes: GHI, DNI y DHI, fundamentales para el análisis energético. La Irradiancia Global Horizontal (GHI) representa la radiación total que llega a una superficie horizontal, combinando tanto la radiación directa como la difusa, por lo que es el indicador más utilizado en estudios fotovoltaicos convencionales. La Irradiancia Directa Normal (DNI) corresponde exclusivamente a la radiación que llega en línea recta desde el sol, medida sobre una superficie perpendicular a los rayos solares; es esencial para sistemas con seguimiento solar y tecnologías de concentración (CSP). Finalmente, la Irradiancia Difusa Horizontal (DHI) es la radiación que llega dispersada por la atmósfera, proveniente de nubes y del cielo, y permite caracterizar la variabilidad debida a la nubosidad. Estas tres variables permiten describir de manera completa el comportamiento de la radiación incidente en un sitio, siendo claves para estimar la producción energética y entrenar modelos predictivos.

Los conjuntos de datos actuales son limitados. La investigación sobre problemas relacionados con la energía solar requiere la creación de conjuntos de datos más sofisticados y de mayor tamaño. Si bien es cierto que varias de estas bases de datos se presentan como herramientas tanto para novatos como para profesionales en el diseño de sistemas de generación fotovoltaica, la verdad es que ninguna hace referencia específica a la localización de centrales fotovoltaicas, como es el caso de nuestro dataset que tiene en cuenta todos los países del mundo, incluye atributos relacionados con diferentes categorías, es de uso general y está disponible para su descarga. Lo creamos desde cero con un conjunto de datos extenso y disponible, que incorpora atributos cualitativos y cuantitativos extraídos de diversas fuentes. Su elaboración fue compleja y costosa en tiempo y esfuerzo, a pesar de la automatización de ciertas fases del proceso de desarrollo por medio de diversas herramientas computacionales.

Aunque estos datos son esenciales para el análisis de viabilidad, la revisión evidenció un predominio de estudios de caso con alcance local, careciendo de una base generalizada a nivel global. La principal diferencia de este conjunto de datos radica en su escala global y robustez, superando las limitaciones de los datasets localizados convencionales. La propuesta se distingue por integrar factores topográficos cualitativos junto con parámetros técnicos críticos, como la inclinación óptima y el potencial fotovoltaico. Si bien existen bases de datos parciales, la singularidad de este trabajo consiste en la consolidación simultanea de variables de ubicación, climáticas, territoriales y técnicas en un único repositorio de alcance mundial, una integración escasa en la literatura actual.

2.2 Modelos de predicción de energía solar

El análisis de la bibliografía resalta el uso de herramientas geoespaciales integradas como métodos de ponderación. La investigación de [7] utilizó un enfoque de GIS combinado con la técnica MIF para identificar ubicaciones óptimas para

PV. La herramienta creó mapas temáticos con factores como radiación solar, temperatura del suelo, humedad relativa, elevación, uso del suelo y distancia a carreteras, asignando pesos a cada valor mediante MIF según una importancia relativa, con la finalidad de identificar ubicaciones óptimas.

En [1] se identificó al GIS como una herramienta fundamental para el análisis espacial, empleando MCDM con criterios de radiación solar, proximidad a redes y otros. Complementándolo con ML resolvieron las limitaciones predictivas del GIS utilizando algoritmos como Random Forest, XGBoost y CNN modelando relaciones no lineales en datos topográficos y meteorológicos, mejorando significativamente la precisión en la identificación de ubicaciones óptimas.

Un estudio en Polonia aplicó un modelo de GIS denominado SILICON con parámetros de pendientes, áreas urbanas, bosques, infraestructura, etc., para evaluar la viabilidad técnica y económica de sistemas fotovoltaicos a gran escala, integrando análisis de idoneidad del terreno con un desglose detallado del coste nivelado de la energía. El modelo identificó que solo el 3.61% del territorio es apto para instalaciones, lo que equivale a un potencial de capacidad anual de 51.4 a 73.4 TWh. La investigación proporciona una herramienta escalable para apoyar la planificación energética y la transición hacia fuentes renovables sin importar las limitaciones del territorio [8].

La investigación en energía solar fotovoltaica se ha centrado principalmente en ubicaciones terrestres; sin embargo, hay estudios centrados en sistemas flotantes, como es el caso del Lago Maint (Filipinas). En este caso se integró IA para el monitoreo del sistema, empleando una metodología descriptiva que combina análisis bibliográfico, datos geoespaciales y trabajo de campo. La característica del recurso mediante GIS determinó, a partir de mapas de irradiancia global, un potencial de generación anual de 762.96 MWh. Para el monitoreo inteligente se definieron componente basados en IA, incluyendo sensores ambientales y eléctricos, microcontroladores y una unidad central de procesamiento para el diagnóstico automático de fallos mediante aprendizaje automático [9].

En el año 2025 en Turquía se determinó la ubicación para nuevas plantas de energía solar en base al máximo potencial de producción utilizando el proceso analítico jerárquico para tomar decisiones, considerando el cálculo de las puntuaciones de impacto para la radiación solar, la temperatura, la humedad, la turbidez y los factores de viento. Su principal contribución es que calculó el impacto de dichas variables y validó los resultados obtenidos comparándolos con la producción real anual y estacional de plantas existentes en tres provincias turcas, lo cual superó a investigaciones previas [10].

En [11] se analizan la aplicación de técnicas de IA, ML y DL, aprendizaje por refuerzo (RL) y lógica difusa en sistemas de energía renovable. El estudio se enfoca en la optimización de la generación energética, la predicción de demanda, el mantenimiento predictivo y la gestión descentralizada. Las variables analizadas abarcan datos meteorológicos, métricas de rendimiento de equipos, parámetros de red eléctrica, consumo energético y variables de mercado. Además, se evalúan casos de implementación real que demuestran mejoras en eficiencia operativa y confiabilidad.

Los modelos de ML que se combinan con herramientas de análisis espacial GIS se han implementado con éxito en la predicción de grados adecuados para la construcción de plantas de energía solar y eólica. Al vincular el inventario de granjas de energía renovable existentes y las variables explicables relacionadas, la metodología basada en ML puede pronosticar con precisión la probabilidad de instalación (idoneidad) en cada unidad geográfica. Los resultados del análisis podrían contribuir al conocimiento sobre los factores clave y sus efectos marginales en la selección de sitios de plantas de energía verde [12].

La aplicación de IA en sistemas fotovoltaicos aborda cuatro problemáticas principales: el seguimiento del punto de máxima potencia (MPPT), la predicción de la producción energética, detección de defectos en módulos FV y la estimación de parámetros de modelos equivalentes. Técnicas como redes neuronales, lógica difusa, metaheurísticas y sistemas híbridos superan el rendimiento de los métodos convencionales, incrementando la eficiencia, precisión y adaptabilidad del sistema. No obstante, su implementación enfrenta limitaciones como la escasez de datasets extensos y los elevados costos computacionales. Pese a estos desafíos, la IA representa un campo de investigación activo y prometedor para la optimización de la producción y el mantenimiento de instalaciones solares [13].

En [14] se presenta un conjunto de datos global de paneles fotovoltaicos para el periodo 2019-2022, con resolución de 20 metros. Para superar las limitaciones de las bases existentes, el estudio implementa un marco de dos etapas que integran aprendizaje profundo y automático. En la primera etapa, un modelo U-Net entrenado con imágenes de alta resolución de Google Earth generando muestras positivas precisas. En la segunda, se aplica un clasificador de aprendizaje positivo no etiquetado con bosques aleatorios (PUL-RF), utilizando dichas muestras para identificar paneles FV a gran escala en imágenes Sentinel-2. Los resultados alcanzan una precisión superior al 90%, proporcionando un detalle espacial sin precedentes para la investigación en el sector de energía solar. La revisión bibliográfica identifica parámetros estándar como coordenadas, elevación, pendiente, curvatura, aspecto y dirección del viento, el aporte distintivo de este estudio radica en la valoración cualitativa de dichas variables. Metodológicamente, se procedió a la digitalización vectorial y cálculo de cada área, aplicando un control de calidad para eliminar duplicados y validar la existencia operativa de las plantas mediante una inspección satelital en ArcMap (Google Earth). Posteriormente, los polígonos fueron

clasificados por su tamaño y sometidos a un análisis de proximidad respecto a la red vial sudamericana. La integración de estos factores permitió determinar la aptitud solar, identificando zonas favorables para la implementación de paneles fotovoltaicos, además de brindar una base robusta de datos.

3 Materiales y métodos

El proyecto se enfoca en la creación de un dataset de calidad sobre el inventario de plantas solares a nivel mundial. Esto constituye un recurso primordial para implementar una aplicación en el contexto de la energía solar: el desarrollo de un modelo de IA basado en deep learning para la predicción de localización óptima de nuevas plantas solares. En esta sección, se detalla la metodología implementada para la creación del dataset, la cual abarca diversas etapas incorporadas en el flujo de trabajo que se representa en la Figura1. El producto final combina exitosamente variables de diversos tipos en un único conjunto de datos coherente, que integra atributos tanto cuantitativos como cualitativos, cuya calidad se basa en el rigor metodológico de las fases de recopilación, limpieza, etiquetado y validación de los datos. Se pone de relieve la importancia de crear registros precisos y actualizados de las instalaciones de energía solar en diversas regiones, especialmente en los mercados emergentes, donde el potencial de la energía solar es enorme.

3.1 Búsqueda bibliográfica

La fase inicial del proyecto consistió en la búsqueda exhaustiva de información bibliográfica con el objetivo de identificar los parámetros determinantes para la localización óptima de plantas fotovoltaicas. se emplearon herramientas de investigación asistida por IA como *STORM* y *Consensus*, las cuales permitieron filtrar la literatura existente a partir de consultas con palabras clave como: "dataset, solar panels, photovoltaic, power, AI". Se excluyeron los sitios web informales, artículos de opinión y estudios sobre otro tipo de energía renovable, independientemente de la coincidencia en sus variables. Se incluyeron únicamente los artículos publicados en fuentes confiables en los últimos 10 años, en los que se especifiquen características del terreno para la localización de proyectos fotovoltaicos. Como resultado, se obtuvieron investigaciones de alto impacto y credibilidad científica, incluyendo revistas indexadas en bases de datos como *Elsevier* y *MDPI*, así como repositorios académicos y de datos como *Google Académico* y *Kaggle*.

3.2 Definición de atributos

La selección de las variables más adecuadas es una tarea de los investigadores de acuerdo con sus necesidades [15]. El propósito de mantener un registro completo de plantas solares fotovoltaicas, que incluya variables tanto explicativas como etiquetas para entrenar un modelo basado en Deep Learning para identificar la ubicación óptima de centrales de generación solar fotovoltaica. El análisis comparativo de los documentos obtenidos de la revisión bibliográfica permitió conocer que se han propuesto diversas variables en la literatura a lo largo del tiempo. La selección de variables se fundamentó en su frecuencia de aparición en las fuentes consultadas, permitiendo la extracción de atributos tanto cualitativos como cuantitativos, los cuales se agrupan en factores geográficos, climáticos y técnicos.

3.3 Estructura del dataset

El conjunto de datos se compone de 27 variables organizadas en cuatro categorías: Identificación, Ubicación, Terreno y Parámetros Técnicos. Estas variables integran atributos de naturaleza tanto cuantitativa como cualitativa Tabla 2.

Las variables de identificación permiten organizar y rastrear cada sitio evaluado, mientras que los atributos de ubicación y terreno (latitud, longitud, elevación, pendiente, aspecto y curvatura) describen las características físicas que determinan la exposición solar y la aptitud del terreno para la instalación. Las variables climáticas (irradiancia, temperatura, humedad y viento) son esenciales para evaluar la disponibilidad de recursos energéticos y las condiciones que afectan la eficiencia de los módulos solares. Finalmente, los atributos técnicos (inclinación y potencia estimada) permiten relacionar el potencial del sitio con la capacidad real de generación. En conjunto, estos atributos permiten construir un dataset robusto y adecuado para comparar alternativas y seleccionar los emplazamientos más favorables para el desarrollo fotovoltaico.

Se incorporó una categorización cualitativa para las variables de terreno (aspecto, pendiente, curvatura) y dirección del viento. Asimismo, se calculó la superficie individual de cada planta solar, categorizándola según su magnitud en: pequeña, mediana y grande. Finalmente, se integró el índice de 'Aptitud Solar' normalizando a una escala cualitativa.

Figure 1: Metodología utilizada para la creación del dataset de plantas solares.

Table 2: Estructura del dataset de plantas solares.

Atributo	Categoría	Tipo	Unidades	Descripción
Código	Identificación	Cualitativo	-	Identificador único alfanumérico de la planta
Nombre	Identificación	Cualitativo	-	Denominación de la planta solar
Estado	Ubicación	Cualitativo	-	Situación actual de la planta solar
País	Ubicación	Cualitativo	-	Lugar nacional de la planta
Longitud	Ubicación	Cuantitativa	Grados	Coordenada geográfica asociada con los meridianos
Latitud	Ubicación	Cuantitativa	Grados	Coordenada geográfica asociada con los paralelos
Area	Terreno	Cuantitativa	m^2	Superficie total de la planta solar
Elevación	Terreno	Cuantitativa	$msnm$	Altura vertical de la planta solar
Pendiente	Terreno	Cuantitativa	Grados	Angulo de inclinación del terreno
Aspecto	Terreno	Cuantitativa	Grados	Orientación de la pendiente del terreno
Curvatura	Terreno	Cuantitativa	-	Forma de la superficie del terreno
Distancia a la vía	Terreno	Cuantitativa	m	Distancia de la planta a la vía más cercana
Tamaño	Terreno	Cualitativa	-	Clasificación del área
T_pendiente	Terreno	Cualitativo	-	Clasificación de la pendiente
T_aspecto	Terreno	Cualitativo	-	Clasificación del aspecto
T_curvatura	Terreno	Cualitativo	-	Clasificación de la curvatura
Irradiancia	Clima	Cuantitativo	kWh/m^2	Cantidad total de energía solar recibida
Temperatura	Clima	Cuantitativo	$\circ C$	Temperatura ambiente promedio del aire en el sitio
Humedad	Clima	Cuantitativo	$\%$	Contenido de vapor de agua en la atmósfera
Viento	Clima	Cuantitativo	m/s	Rapidez del flujo de aire en el sitio
Dirección	Clima	Cuantitativo	Grados	Punto cardinal desde donde sopla el viento dominante
Aptitud solar	Clima	Cuantitativa	-	Indice solar
DT_Viento	Clima	Cualitativo	-	Clasificación de la dirección del viento
T_aptitud	Clima	Cualitativo	-	Clasificación de la aptitud solar
Inclinación	Operacional	Cuantitativo	Grados	Ángulo ideal del panel
Potencia	Operacional	Cuantitativo	kWp	Máxima potencia que el sistema solar puede generar
Capacidad	Operacional	Cuantitativo	kW	Capacidad total estimada por el sistema

3.4 Selección de fuentes de datos

Una vez determinadas las variables que conforman el dataset, en esta sección se explica de dónde y cómo se obtuvieron los datos: las fuentes primarias, el proceso de adquisición y las técnicas empleadas para recolectar información precisa de las plantas solares fotovoltaicas.

Los datos fueron adquiridos de fuentes en línea públicas, confiables, actualizadas, de acceso abierto y gratuito, en las cuales es posible consultar, descargar, reutilizar y compartir, útil para la investigación, planificación y toma de decisiones.

Desde la aparición de la Web en la década de 1990, el volumen de información disponible abarca todos los temas [16, 17]. En esta plataforma mundial, la información sobre plantas solares, recursos energéticos y condiciones geoespaciales se ha incrementado de manera exponencial.

Encontramos los siguientes recursos fundamentales para la obtención de los datos: Global Solar Atlas, Global Energy Monitor, Global Solar Power Tracker y NASA POWER.

3.4.1 Global Solar Atlas

La plataforma web de Global Solar Atlas (GSA) es una herramienta de acceso gratuito en línea desarrollada por el Grupo del Banco Mundial y la empresa Solargis. Proporciona datos globales de irradiación solar y potencial fotovoltaico (FV) a nivel mundial. Es un recurso clave para la planificación energética, la investigación y, especialmente, para la localización óptima de plantas solares. El sitio web ofrece un mapa interactivo mundial que permite ver datos solares (GHI, DNI, PVOUT) a alta resolución para cualquier punto del planeta. Se pueden descargar capas de datos GIS para el análisis en software especializado (como QGIS o ArcGIS). El Atlas considera irradiación solar, temperatura, inclinación óptima, elevación, etc., pero no incluye (por defecto) factores locales como sombras por vegetación, edificaciones, sombras temporales, calidad del suelo, acceso a red eléctrica, regulaciones locales, etc.

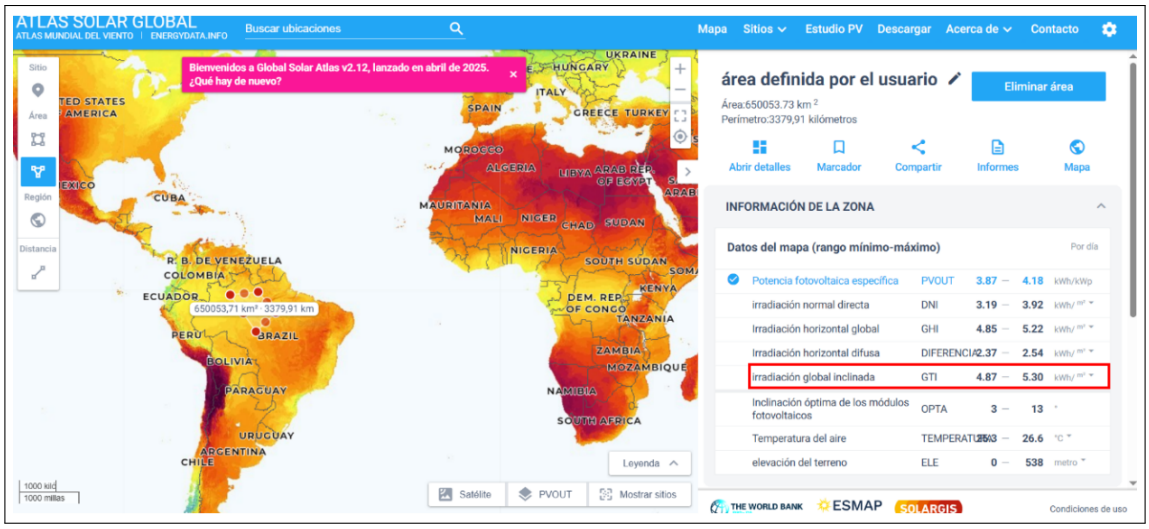


Figure 2: Página web del Atlas Solar.

3.4.2 Global Energy Monitor

El sitio Web de la organización Global Energy Monitor (GEM) [1] es una referencia global. Documenta y analiza la infraestructura energética mundial, tanto de combustibles fósiles como de energías renovables. Ofrece de manera gratuita bases de datos públicas a través de rastreadores ("trackers") globales que recopilan información sobre proyectos de energía en todo el mundo. A través de su plataforma y herramientas interactivas, es posible visualizar la distribución geográfica de infraestructuras energéticas, su estado (activa, planificada, retirada), la capacidad instalada y otros datos clave. Los datos son accesibles para su descarga y análisis

Uno de los rastreadores relevantes de GEM es el Global Solar Power Tracker (Rastreador de Energía Solar Global)[2], el cual es un inventario que cubre proyectos operativos y planificados de plantas solares.

3.4.3 NASA POWER

El proyecto NASA Prediction Of Worldwide Energy Resources (POWER) [3] es una iniciativa de la NASA proporciona acceso gratuito a datos meteorológicos y de recursos solares. NASA POWER es una base de datos robusta, de alta calidad científica y gratuita que proporciona la materia prima (datos climáticos y solares) necesaria para los análisis detallados de viabilidad y localización de proyectos de energía solar y otras renovables. La NASA facilita el acceso a estos datos relativos a irradiación solar y variables climáticas/meteorológicas. Interfaz web via un visor interactivo llamado POWER Data Access Viewer (DAV), que permite seleccionar ubicación geográfica (latitud/longitud), periodo de tiempo, variables de interés, y visualizar los datos en mapas o gráficos. con una interfaz web interactiva que permite a los usuarios seleccionar una ubicación específica (coordenadas) o una región, elegir la comunidad de interés (Ej: Renewable Energy), seleccionar los parámetros deseados y descargar los datos en formatos estándar como CSV. En la interfaz → tu puedes indicar un punto geográfico (latitud/longitud), seleccionar los parámetros que necesitas (irradiancia, temperatura, viento, etc.), elegir la escala temporal (hora, día, mes, año, climatología) y descargar los datos.

3.5 Adquisición de datos

Los datos requeridos pertenecen a categorías de identificación nominal, ubicación geográfica, características del terreno, climáticas y de carácter técnico.

3.5.1 Datos de identificación

A partir de Global Solar Atlas, se obtuvieron los datos relacionados con la identificación de las plantas solares. Esto comprende la definición de un *código* alfanumérico como un identificador único de cada planta solar, el cual se genera

[1] `https://globalenergymonitor.org/`
[2] `https://globalenergymonitor.org/projects/global-solar-power-tracker/`
[3] `https://power.larc.nasa.gov/`

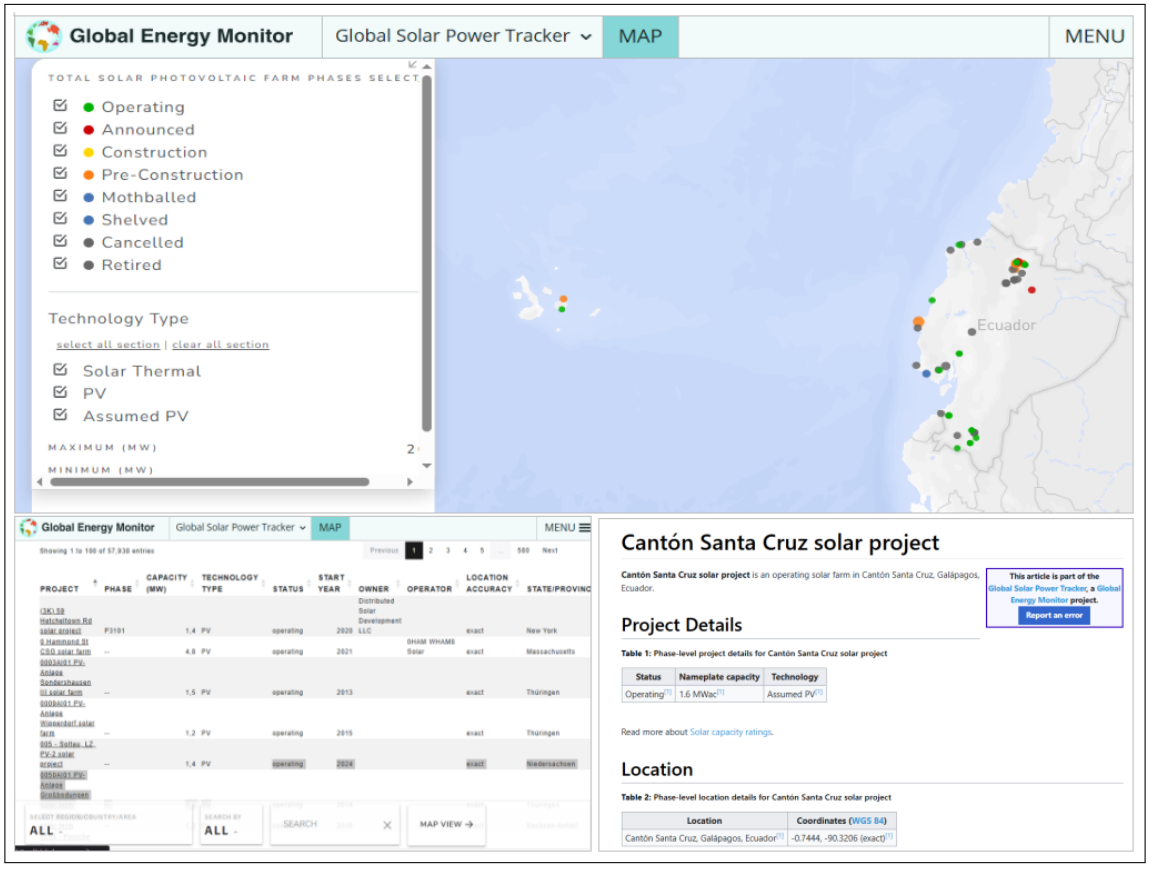


Figure 3: Página web de Global Energy Monitor.

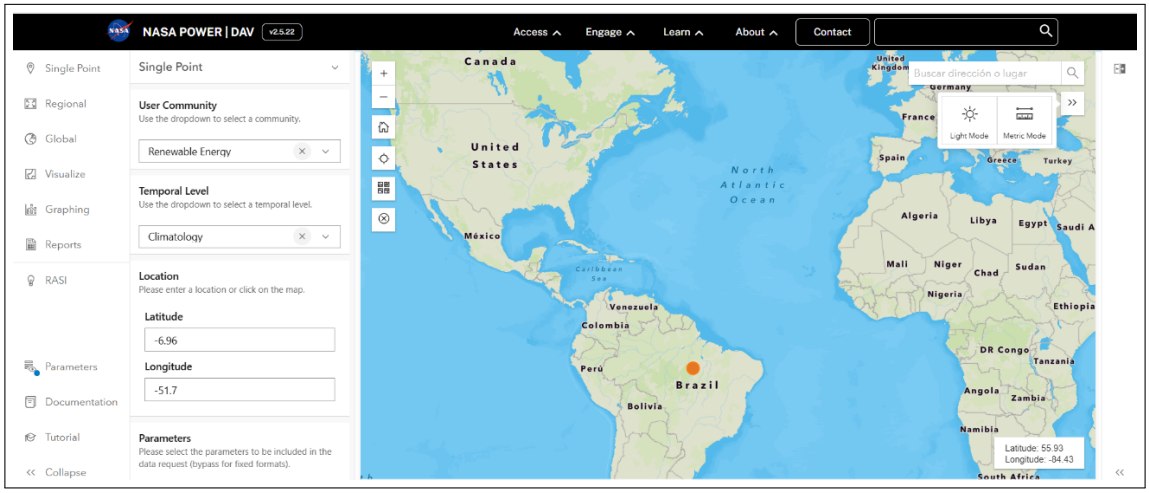


Figure 4: Página web de Nasa Power.

mediante la concatenación de tres elementos: el identificador original de la plataforma Global Solar Atlas, las tres primeras letras del país de origen y la inicial de la categoría de tamaño de la planta (P,M,G). Por otra parte, el *nombre* asignado a la planta solar corresponde a la designación oficial del proyecto fotovoltaico según consta en la fuente primaria. Por ejemplo: la planta solar denominada "Cantón Santa Cruz solar project", ubicada en las Islas Galápagos (Ecuador), se identifica con el código "37402-ECU-P".

3.5.2 Datos de ubicación geográfica

Desde el portal de Global Energy Monitor, opción Global Solar Power Tracker, se descarga el inventario de plantas solares. Previamente, se debe completar un formulario de solicitud básico (nombre, correo institucional y propósito del uso). Una vez aprobado el acceso, se obtienen los archivos en formatos CSV, Shapefile (SHP) y GeoJSON. Estos datos se cargan en un GIS: la capa de puntos (SHP y GeoJSON) mediante "Añadir capa vectorial", mientras que los archivos CSV se importaron como "Texto delimitado". Las columnas de *latitud* y *longitud* son trabajadas en el sistema de referencia espacial "WGS 84 (EPSG:4326)" para asegurar su correcta ubicación. Se realizó la verificación de la georreferenciación y consistencia espacial de los puntos. Finalmente, la capa procesada se exportó en el formato requerido (SHP, GeoJSON o GPKG), asegurando su compatibilidad con los análisis espaciales.

Primero, cada punto fue superpuesto sobre las capas de irradiancia, aplicando el paso "comparar la ubicación del punto con el recurso solar disponible", lo que permitió identificar registros ubicados fuera de zonas compatibles con generación fotovoltaica. Este contraste inicial facilitó detectar coordenadas erróneas o inconsistentes con la distribución real del recurso solar. Luego, en ArcGIS se efectuó una revisión detallada mediante "inspección visual georreferenciada", "confirmación mediante imágenes satelitales" y "verificación de infraestructura observable", asegurando que cada punto correspondiera a una instalación solar identificable en imágenes de alta resolución. Asimismo, se generaron buffers y consultas espaciales para "evaluar coherencia espacial con uso de suelo y límites geográficos", lo que permitió detectar desplazamientos o referencias incorrectas en la geolocalización. Para eliminar sesgos se implementaron tres procedimientos principales: "detectar y eliminar duplicados", mediante análisis de proximidad y revisión de atributos; "identificar outliers espaciales", utilizando métricas de distribución como Average Nearest Neighbor; y "corregir puntos sobrepuestos o mal diferenciados", revisando registros con coordenadas idénticas o solapadas. Estos pasos permitieron depurar el dataset y asegurar que la capa utilizada en los análisis representara únicamente instalaciones reales y correctamente posicionadas.

Para la ubicación geográfica de los paneles solares existentes a nivel mundial, se integraron múltiples fuentes de información. La principal referencia fue el Atlas Solar y Global Energy Monitor, ya que concentran la mayor densidad de instalaciones fotovoltaicas registradas y constituyen la base más sólida para este tipo de análisis. Además, se incorporaron datos provenientes de imagenes raster, de las cuales se extrajeron los atributos relevantes para el estudio.

Todos los conjuntos de datos fueron armonizados en un único dataset, para lo cual se estandarizó el sistema de coordenadas a WGS 84 (auxiliar) y, cuando fue necesario, se reproyectaron las capas originales. Este proceso garantizó la compatibilidad espacial y permitió disponer de una base de datos unificada y coherente para el análisis posterior. La verificación espacial de los puntos de ubicación se realizó empleando un basemap integrado (mapas satelitales) de ArcGIS como referencia cartográfica. Posteriormente, las áreas de cada instalación fueron delimitadas mediante un proceso de digitalización manual (poligonización). El cálculo de las superficies se efectuó utilizando la Calculadora Geométrica de ArcMap, lo que permitió cuantificar con precisión las dimensiones de cada polígono y corroborar la exactitud de su localización geográfica. Este procedimiento aseguró la consistencia geométrica y espacial de las plantas dentro del conjunto de análisis.

3.5.3 Datos del terreno

Las variables del terreno son críticas para determinar la localización óptima de plantas solares Se derivan matemáticamente a partir de un Modelo Digital de Elevación (DEM). Un DEM es una representación tridimensional de la superficie terrestre en forma de una cuadrícula regular basada en píxeles, donde cada celda o píxel contiene un valor de altura. Todas las variables generadas a partir del DEM mantienen ese mismo formato matricial, por lo que deben ser tratadas como mapas raster.

Adquisición y preparación del DEM

El modelo de elevación digital (DEM) proviene del "Global Solar Atlas (GSA)", descargados directamente desde su plataforma oficial. El área de cobertura abarca la totalidad de Sudamérica. Sin embargo, previo a la utilización del DEM, la preparación del DEM es un proceso fundamental para garantizar resultados confiables. Este proceso consta de 7 pasos principales, cuyo tratamiento se realiza en ArcMap.

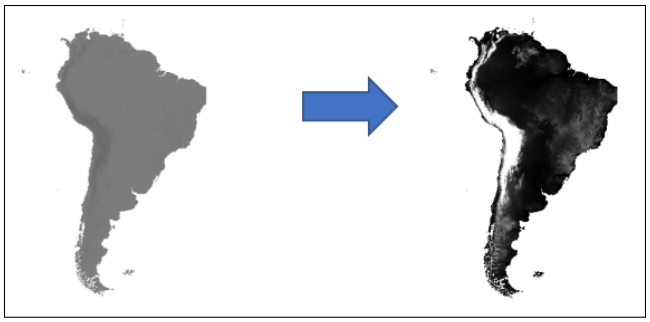


Figure 5: Correccion Fill.

1. Verificación del formato y compatibilidad: Se encuentra en formato raster GeoTIFF, compatible con los entornos SIG y resolución espacial de 9 arc-segundos es apropiada para los objetivos del estudio.
2. Revisar el sistema de coordenadas: Se debe verificar el sistema de referencia espacial del DEM. Los análisis geomorfológicos incluyendo pendiente, aspecto y curvatura requieren un sistema de coordenadas proyectado, ya que estos cálculos deben realizarse en unidades métricas y no angulares. La reproyección de grados a metros (sistema UTM basado en WGS84) se realizó mediante la herramienta Project Raster disponible en ArcGIS
3. Comprobar la integridad del DEM con el fin de identificar posibles inconsistencias espaciales. Esta inspección incluye la revisión visual de patrones irregulares, como superficies artificialmente planas, líneas o bloques sin datos y variaciones abruptas de elevación que no corresponden a la morfología real del terreno. Adicionalmente, se verifica el valor asignado a NoData mediante la pestaña Source del raster. En caso de ausencia de una máscara adecuada, se genera una máscara explícita utilizando herramientas como Set Null o mediante el editor de ráster, asegurando así una delimitación correcta de las áreas válidas
4. Corregir depresiones o picos artificiales mediante la herramienta Fill (rellenar depresiones), ya que estos errores pueden interrumpir el flujo de datos y generar resultados incorrectos (Spatial Analyst → Hydrology → Fill).
5. Suavizar si es necesario, aplicando un suavizado moderado: Filter (Low Pass) y Focal Statistics (media, mediana).
6. Ajustar el Z-Factor para asegurar coherencia entre las unidades horizontales y verticales, normalmente fijándolo en 1 si se trabaja en metros. El Z-factor afecta al cálculo de Pendiente, Curvatura, Aspecto y Sombreado.
7. Finalmente, el DEM debe recortarse (Clip Raster) al área de estudio para mejorar el rendimiento y asegurar coherencia espacial con las demás capas (Data Management → Raster → Raster Processing → Clip). Asegurar una resolución consistente con múltiples capas: Resample si es necesario.

Siguiendo estos pasos, se obtiene un DEM limpio, continuo, correctamente proyectado y listo para generar mapas de pendiente, aspecto, curvaturas confiables.

Generación de mapas ráster

El insumo necesario para la generación de los mapas ráster es el Modelo Digital de Elevación (DEM) en formato raster. Todos los mapas de las variables geomorfológicas se derivan del DEM. Por tanto, todos estos mapas deben tener la misma resolución de píxel. A partir del DEM preparado, automáticamente podemos obtener varios mapas de distintas variables. La generación de estos mapas se basa en el "análisis del relieve" utilizando herramientas del módulo "Spatial Analyst".

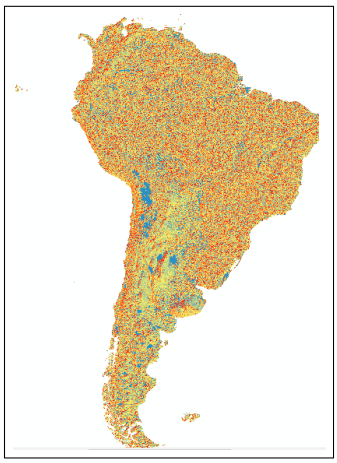
Mapa de sombras (Hillshade): resalta las formas del terreno y facilita su interpretación visual. Permite visualizar la iluminación del terreno simulando la posición del sol para resaltar montañas, valles y formas principales del relieve. Se genera a partir del DEM reprocesado a 100 × 100 m mediante la herramienta de sombreado disponible en ArcGIS dentro de las funciones de análisis de superficie. Se utilizan parámetros comunes de iluminación (azimuth 315° y altitud 45°) y se activa el modelado de sombras para dar un efecto más realista. El factor Z es 1 cuando el DEM está en metros.

Mapa de pendiente: Representa la inclinación del terreno expresada en grados o porcentaje, derivada directamente del DEM mediante la herramienta de cálculo de pendiente incluida en las funciones de superficie de ArcGIS. Se utiliza un factor Z igual a 1 cuando la elevación está en metros. Posteriormente, la pendiente puede clasificarse en rangos

(a) Mapa de sombras.

(b) Mapa de pendientes.

(c) Mapa de curvatura.



(d) Mapa de aspecto.

Figure 6: Mapas rásters de variables de terreno a partir del DEM.

desde las propiedades del raster para facilitar su interpretación. Este mapa es fundamental en estudios geomorfológicos, modelación de drenaje, evaluación de riesgos y análisis de aptitud territorial.

Mapa de curvatura: Permite identificar concavidades y convexidades del terreno mediante el cálculo de la curvatura a partir del DEM utilizando las funciones de superficie de ArcGIS (curvature). Este análisis emplea un factor Z igual a 1 cuando la elevación está en metros. Los valores positivos representan superficies convexas asociadas a crestas o divisorias; los valores negativos corresponden a superficies cóncavas vinculadas a zonas de acumulación o valles; y los valores cercanos a cero indican áreas planas o de transición. Este mapa es fundamental para evaluar la dinámica del flujo superficial, procesos de erosión, estabilidad del terreno y análisis geomorfológico detallado.

Mapa de aspecto: Determina la orientación de las laderas a partir del DEM mediante la herramienta de aspect disponible en las funciones de superficie de ArcGIS, empleando un factor Z igual a 1 cuando la elevación está expresada en metros. El resultado expresa la dirección de máxima pendiente en valores de 0° a 360°, asignando el valor –1 a las zonas completamente planas. Este mapa identifica la orientación cardinal del relieve (N, NE, E, SE, S, SW, W, NW) y es esencial en estudios climáticos, forestales, agrícolas y de modelación ambiental, donde la exposición al sol y al viento influye directamente en los procesos del territorio.

Integración de mapas mediante algebra map

Una vez generados los mapas, se pueden combinar o utilizar en análisis posteriores:

Hillshade + Pendiente: mejora visual del relieve.

Para mejorar la visualización del relieve, se puede combinar el mapa de sombras (hillshade) con el mapa de pendientes mediante álgebra de mapas usando Raster Calculator. Una expresión común es ("Hillshade" * 0.7) + ("Slope" * 0.3), donde el hillshade aporta el efecto de iluminación y la pendiente incrementa el contraste topográfico. Esta combinación permite resaltar formas del relieve que no se aprecian claramente cuando se visualizan por separado, creando un producto más expresivo y útil para cartografía y análisis geomorfológico.

Aspect + Slope: análisis solar y microclimas.

Para estudiar la interacción entre la orientación del terreno (aspect) y la inclinación (slope), se puede realizar álgebra de mapas normalizando ambas capas para que sean comparables. Una expresión típica es ("Aspect" / 360) * ("Slope" / max_slope), lo que permite obtener un índice combinado de exposición. Este resultado es útil en estudios de microclima, energía solar, vegetación o distribución de humedad, ya que integra tanto la dirección de exposición solar como la intensidad del ángulo de la ladera.

Curvature + Slope: erosión y estabilidad de laderas.

Para analizar procesos erosivos y estabilidad del terreno, una forma simple de combinar curvatura y pendiente es sumar ambas capas, por ejemplo, mediante ("Curvature") + ("Slope"). También puede usarse una versión ponderada ("Curvature" * 0.5) + ("Slope" * 0.5) si se quiere equilibrar su influencia. La curvatura indica si la superficie es cóncava o convexa, mientras que la pendiente representa la inclinación del terreno; combinarlas ayuda a identificar zonas potencialmente inestables o con mayor susceptibilidad a erosión o acumulación de materiales.

Aspect + Hillshade: modelación de iluminación para energía solar.

Para integrar orientación y nivel de iluminación en un solo índice enfocado en energía solar o modelación térmica, se puede usar un álgebra de mapas que normalice ambas capas. Una expresión común es ("Hillshade" / 255) * ("Aspect" / 360). Esta operación combina la cantidad de luz recibida según el hillshade con la dirección hacia la que se orienta cada ladera mediante el aspect. El resultado permite identificar zonas mejor iluminadas y con orientación favorable para la instalación de paneles solares o estudios de confort climático.

Los mapas de pendiente, aspecto, sombras y curvatura, derivados de un DEM, son herramientas fundamentales para determinar zonas con potencial fotovoltaico, porque permiten evaluar de manera precisa las condiciones topográficas que influyen directamente en la captación de energía solar. Cada mapa aporta información complementaria, y cuando se integran, permiten identificar sitios óptimos, seguros y eficientes para la instalación de paneles solares.

Mapa de aptitud Solar

Un mapa de aptitud solar es un producto cartográfico generado mediante análisis geoespacial que integra múltiples variables físicas y ambientales —como la irradiancia solar, pendiente, orientación del terreno (aspecto), sombras, curvatura y restricciones territoriales— para evaluar la idoneidad de un área para la instalación de sistemas de energía solar. A partir de estas variables se calcula un índice de aptitud que clasifica el territorio en zonas de mayor o menor favorabilidad, permitiendo identificar de forma objetiva los sitios con mejores condiciones para el aprovechamiento del recurso solar, apoyar la planificación energética y optimizar la toma de decisiones en proyectos fotovoltaicos o termosolares.

El proceso general para la generación del mapa de aptitud solar incluye:

1. Integración de datos al entorno SIG: las capas base necesarias: irradiancia solar, sombras, pendiente, curvatura y aspecto. Todas las capas se estandarizan al mismo sistema de referencia y resolución espacial.
2. Reclasificación de variables topográficas: cada capa temática se reclasifica según criterios de aptitud: las pendientes suaves, orientaciones favorables (aspecto), curvaturas estables y zonas sin sombras reciben valores más altos de aptitud.
3. Normalización en una escala común: los valores reclasificados se ajustan a una escala uniforme (por ejemplo, 1–5), permitiendo la comparación y combinación entre variables heterogéneas.
4. Ponderación (opcional): las variables pueden recibir pesos diferenciados según su importancia relativa en la instalación de sistemas solares, utilizando métodos como AHP o juicio experto.
5. Aplicación de álgebra de mapas: las capas normalizadas y, en caso de aplicarse, ponderadas, se integran mediante una operación de suma o superposición ponderada para generar el índice de aptitud solar.
6. Clasificación del resultado: el índice final se clasifica en categorías de aptitud (baja, media, alta, muy alta), facilitando la interpretación espacial y la identificación de áreas favorables.
7. Validación y verificación: el mapa final se revisa para asegurar coherencia espacial y consistencia con patrones reales de radiación solar y características del terreno.

8. Exportación del producto final: el mapa de aptitud solar se guarda en formato GeoTIFF o en el formato requerido, asegurando su compatibilidad con análisis posteriores.

Área y tamaño

La determinación del área se realizó mediante la digitalización manual de polígonos mediante el GIS. Utilizando imágenes satelitales (mapa base) como referencia se vectorizó el perímetro de las plantas en estado operativo. Para los proyectos en fase de construcción o sin visibilidad satelital, se asignó un valor de 30 m^2, con el fin de evitar la exclusión de registros en la base de datos.

Clasificación por tamaño: a partir de las superficies calculadas, se las dividió en tres categorías definidas empíricamente según la muestra operativa verificada: Pequeña (área $< 20.000\ m^2$), mediana (área entre $20.000\ m^2$ y $200.000\ m^2$) y grande (área $> 200.000\ m^2$).

Distancia a la vía

El cálculo de la distancia desde las plantas fotovoltaicas hacia la red vial principal se ejecutó de la siguiente manera:

1. Adquisición de datos vectoriales: Descarga de la cartografía vial de Sudamérica desde el repositorio GEOFABRIK (OpenStreetMap Data Extracts).
2. Preprocesamiento y filtrado (QGIS): Depuración de la base de datos mediante consultas por atributos (SQL), seleccionando exclusivamente vías de primer y segundo orden (primary y secondary).
3. Fusión de capas: Unificación de los archivos vectoriales resultantes en un único shapefile denominado VIAS_UNIDAS.
4. Integración de datos: Carga de a capa VIAS_UNIDAS y del inventario de plantas solares proveniente del Global Solar Atlas.
5. Corrección Topológica: Ejecución del algoritmo "Corregir geometrías" sobre la red vial para subsanar errores de digitalización.
6. Estandarización de Proyección: Reproyeccion de ambas capas al sistema de referencia "South America Albers Equal Area (EPSG:102033)", garantizando una proyección métrica a nivel continental.
7. Análisis espacial: Ejecución de la herramienta ''Unir atributos por el más cercano" con los siguientes parámetros:
   (a) Capas de entrada: Plantas solares (reproyectadas) y Red vial (reproyectadas)
   (b) Vecinos más cercanos: 1
   (c) Resultados: Generación del vector final BASE_FINAL_SUDAMERICA, conteniendo la distancia en metros.

3.5.4 Datos climáticos

Antes de instalar paneles solares, es fundamental evaluar los riesgos medioambientales y las amenazas climáticas físicas que puedan afectar a las operaciones. Las estaciones meteorológicas y los sensores pueden proporcionar datos valiosos para evaluar el rendimiento del sistema y minimizar las pérdidas, garantizando así que los datos recopilados sean relevantes y útiles. Para completar la información geoespacial, se integraron datos climatológicos de alta precisión obtenidos de fuentes especializadas como NASA POWER y el Atlas Solar Global.

En cuanto a la Irradiancia Global se la obtiene de la página Atlas Solar Global, aquí se seleccionará el área de interés, después de generarán datos de Irradiancia normal directa, horizontal global, horizontal difusa y global inclinada, para el estudio se utilizó el último dato (Figura 3).

En la página de NASA POWER (Figura 4), a través de la opción de Single Point, User Community, Renewable Energy, Temporal Level, Climatology, en Location se debe colocar los datos de latitud y longitud, se obtienen los parámetros de humedad relativa, dirección y velocidad del viento. Todos los datos se descargarán en formato CSV, los cuales originalmente se presentan con una distribución mensual y un promedio anual. Para este análisis se optó por utilizar el valor del promedio anual.

3.5.5 Datos operacionales

En los inventarios del Global Energy Monitor, la *capacidad* (MW) registrada de las plantas de generación fotovoltaica se obtiene mediante un proceso sistemático de recolección y validación de datos públicos. Los investigadores compilan información de fuentes oficiales (como datos gubernamentales, permisos y reportes de empresas generadoras), informes

industriales, y noticias verificadas de medios, que luego se contrastan con otras bases de datos públicas y privadas para asegurar coherencia y calidad. Cada instalación documentada tiene una página de perfil con referencias detalladas (GEM.wiki), y todas las capacidades nominales (en megavatios) reflejan la potencia declarada de las unidades de generación bajo condiciones estándar.

La *inclinación óptima* corresponde al ángulo de montaje que maximiza la producción de un sistema fotovoltaico fijo orientado hacia el ecuador. En el Atlas Solar, este parámetro se estima mediante algoritmos derivados del modelo solar de Solargis, los cuales evalúan la respuesta energética del sistema para distintos ángulos de inclinación. Posteriormente, se selecciona el ángulo que proporciona el mayor rendimiento energético en un periodo comprendido entre los años 2018-2024 de acuerdo a la informacion disponible en la pagina oficial. Dicha información se la representa en imágenes ráster las cuales se colocan en un SIG y se realiza la extracción manual del valor del pixel que contiene el dato de la inclinación óptima.

El *potencial fotovoltaico* representa la generación eléctrica específica (kWh/kWp) estimada para un sistema fotovoltaico estándar. Para su cálculo, el Atlas Solar combina la irradiación global inclinada generando un resultado, el cual es un indicador integrado del desempeño energético (2018 al 2024), útil para estudios de factibilidad y planificación energética a escala regional y nacional.

3.6 Compilación y depuración de datos

Finalmente, el proceso concluyó con la integración y consolidación de todos los datos generados previamente en el SIG (ArcMap) junto con los datos climatológicos obtenidos de fuentes internacionales (NASA POWER y Atlas Solar Global), formando una dataset completa y consistente.

Las capas asociadas al recurso solar, entre ellas la "irradiancia global", "directa" y "difusa", poseen una "resolución nominal de 9 arc-segundos"; y para mantener una estructura espacial uniforme en todo el conjunto, todas las capas complementarias (potencial fotovoltaico, elevación del terreno y parámetros climáticos) fueron "estandarizadas mediante resampling" a una resolución común de 9 arc-segundos. Este procedimiento permitió construir una "base cartográfica homogénea" y adecuada para el "modelamiento de aptitud" y la "comparación directa entre variables".

La depuración del dataset se ejecutó mediante un procedimiento sistemático. En primera instancia, se realizó una "validación cruzada" entre el Atlas Solar, Global Energy Monitor y una verificación manual en ArcGIS y QGIS, con el objetivo de identificar inconsistencias en la ubicación de las plantas solares. Este proceso permitió detectar "puntos duplicados", "ubicaciones aproximadas" y "registros sobrepuestos", los cuales podían generar valores repetidos o distorsionados durante la extracción de información geoespacial.

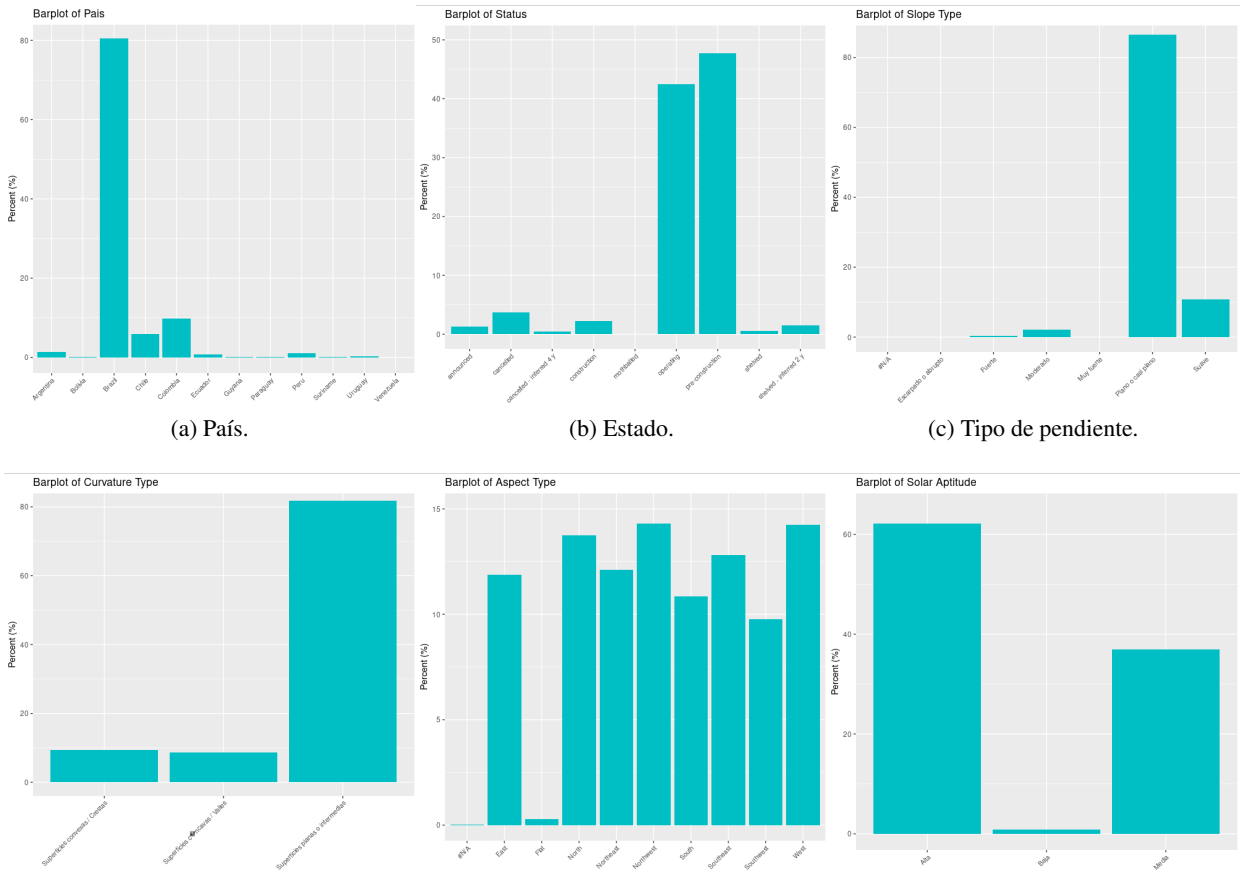
Posteriormente, se evaluó la "coherencia espacial" del conjunto de datos para identificar valores atípicos (outliers). Se eliminaron aquellos puntos ubicados "fuera del área de cobertura de los rásteres", debido a que producían valores extraídos incoherentes o no representativos. Se obtuvo una versión depurada del dataset, garantizando un nivel adecuado de "consistencia", "precisión espacial" y "calidad de la información", requerido para los análisis posteriores en el SIG.

El conjunto de datos inicial en el Atlas Solar comprendía un total de 7006 datos, correspondientes a instalaciones fotovoltaicas en Sudamérica. Se procedió a una evaluación de calidad de los datos mediante una depuración de duplicados utilizando QGIS. La detección de este problema espacial se basó en la coincidencia de los atributos de latitud y longitud. Este proceso fue necesario debido a que durante la fase de digitalización se identificaron múltiples registros para una misma ubicación física de paneles, lo que generaba una redundancia en la dataset. Por ejemplo, en la dataset inicial en Brasil, la instalación con OBJECTID 1054 presentaba una ocurrencia cuádruple con idénticas coordenadas geográficas y parámetros descriptivos. La limpieza de los datos se implementó mediante una herramienta de "detección de duplicados por atributos", resultando en un conjunto consolidado de 5078 registros únicos. La corrección de estos casos es representativa de la metodología aplicada.

4 Análisis Estadístico del Dataset

Esta sección presenta un análisis estadístico básico del dataset, lo que demuestra la utilidad científica de los datos. Se estudian las características más importantes de los atributos que conforman el dataset. Por un lado, se analizan las variables de tipo cualitativo a través de la distribución de las clases. Por otro lado, se analizan las variables de tipo cuantitativo a través de los diagramas de distribución de valores y los indicadores estadísticos más importantes. Estas actividades han sido generadas mediante el lenguaje de programación R y el entorno RStudio. El código implementado está disponible en línea.

4.1 Variables cualitativas



(a) País. (b) Estado. (c) Tipo de pendiente.

(d) Curvatura. (e) Tipo de aspecto. (f) Tipo de aptitud solar.

Figure 7: Diagramas de barras de las variables cualitativas del dataset.

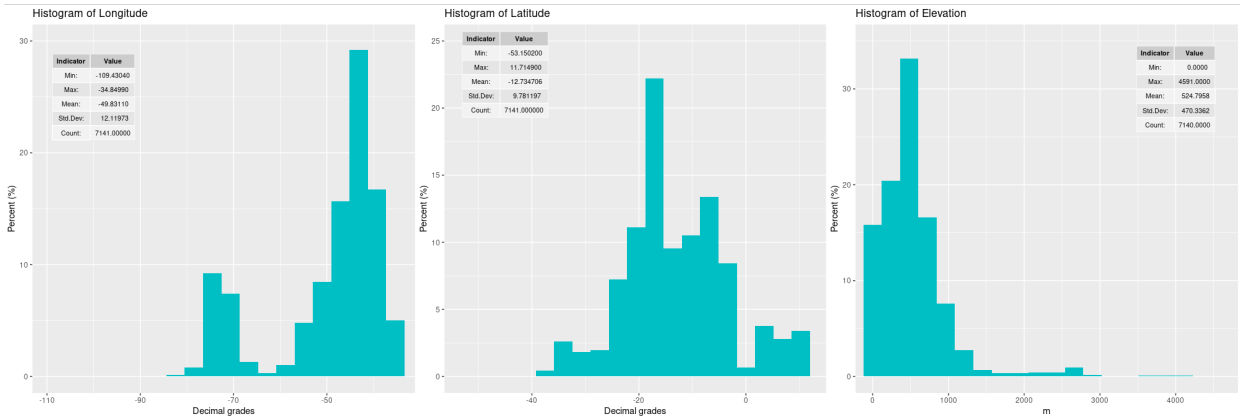
4.2 Variables cuantitativas

5 Caso de uso

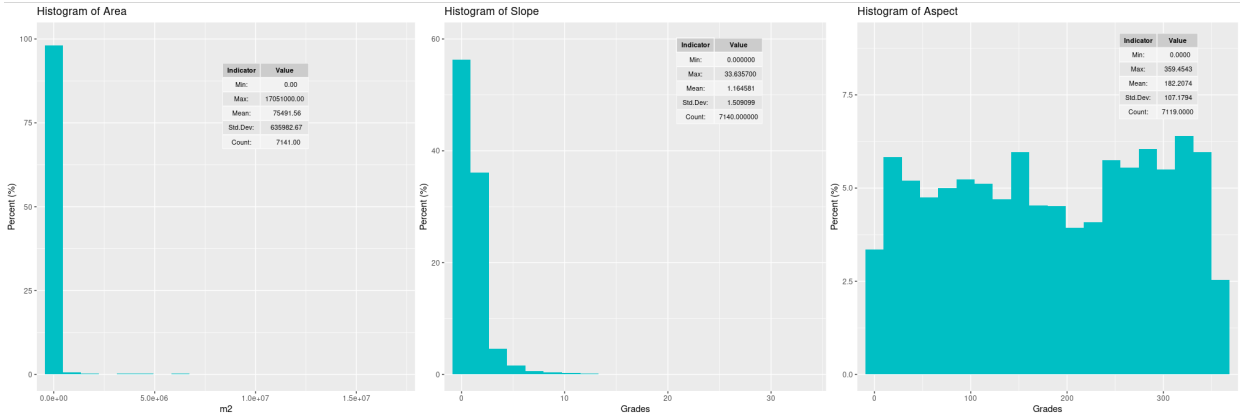
Para demostrar la utilidad y el valor práctico del conjunto de datos de plantas solares desarrollado, la presente sección se centra en un caso de uso aplicado a un problema crítico dentro del sector energético. Por ello, seleccionamos la tarea de localización óptima de plantas solares para ilustrar la capacidad y riqueza de los datos recopilados, lo que puede mejorar significativamente la precisión y la toma de decisiones en el despliegue de nueva infraestructura energética. Para abordar este complejo desafío, hemos optado por un enfoque basado en Machine Learning (ML), el cual permite aprender patrones ocultos en los datos y generar predicciones robustas y confiables. A continuación, se detalla la metodología empleada, que se representa de manera gráfica en la Figura.

5.1 Definición de la estrategia

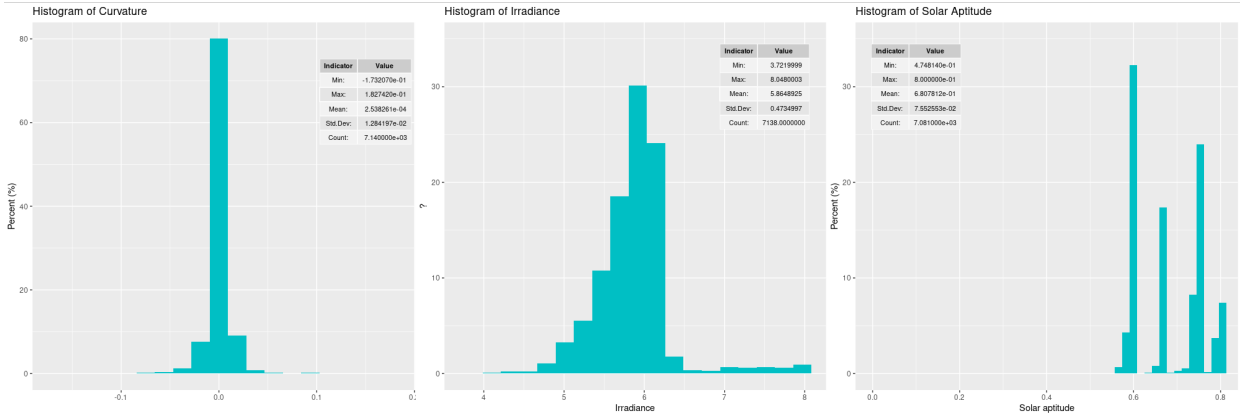
El objetivo es obtener la mejor posición geográfica para ubicar una planta solar maximizando la generación de energía; es decir, inferir la producción estimada mediante el modelo ML. Ya que el dataset contiene la energía real generada por plantas existentes, se puede entrenar y evaluar un modelo base (baseline) de tipo supervisado, lo que es ideal para redes neuronales. Esto convierte al modelo en un predictor de producción, que luego se aplica a lugares donde aún no existen plantas. Se convierte en una herramienta muy valiosa para "pre-seleccionar" zonas con buen recurso solar o realizar estudios de factibilidad.



(a) Longitud (b) Latitud (c) Elevación



(d) Area (e) Pendiente (f) Aspecto



(g) Curvatura (h) Irradiancia (i) Aptitud solar

Figure 8: Histogramas de las variables cuantitativas del dataset.

5.2 Ingeniería de características

La elección de la mejor ubicación geográfica basada en la generación de energía solar se puede enmarcar dentro de un problema de regresión. Dado un conjunto de características geográficas, geomorfológicas, climáticas y operativas, es posible estimar cuánta energía produce una planta solar en ese lugar. Se definen las variables independientes y dependiente. Sin embargo, se debe incluir solo las variables relevantes, descartando aquellas que pueden ser redundantes, irrelevantes o que causan ruido al modelo. La Tabla 3 resume las variables que intervendrán en el entrenamiento del modelo y su respectivo rol.

Table 3: Definición de variables independientes y dependiente.

Variable	**Representación**	**Categoría**
Latitud, longitud, elevación, pendiente, aspecto, curvatura, distancia, temperatura, irradiancia, humedad, velocidad del viento, dirección del viento, inclinación	x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}	Independientes
Potencial fotovoltaico	$\hat{y}$	Dependiente

Las variables consideradas influyen de forma significativa en la cantidad de energía solar disponible y la viabilidad del sitio. Estas variables son imprescindibles porque describen la física, la geografía y el clima operativo donde funcionará la planta. La red neuronal necesita esta información para aprender relaciones espaciales complejas y predecir correctamente la energía anual neta estimada en kWh por kWp y la ubicación óptima para instalar plantas solares.

Por otra parte, las variables de código, nombre, estado, país de la planta solar se descartan porque son atributos que describen identidad administrativa de la planta existente, no tienen relación física con la idoneidad del sitio. Las variables de área, tamaño y capacidad dependen del diseño de la planta; es decir, describen el proyecto más no reflejan las condiciones naturales del terreno. El tipo de pendiente, tipo de aspecto, tipo de curvatura y tipo de dirección del viento son categorías creadas a partir de variables numéricas ya presentes, por lo que su inclusión generaría colinealidad y redundancia. Finalmente, la aptitud solar y el tipo de aptitud solar son variables derivadas que resultan de combinar matemáticamente otras variables, por lo que generarían una dependencia circular. Las variables descartadas no aportan información útil y afectarían la capacidad predictiva del modelo. Se conservan únicamente las variables físicas, geográficas y climáticas primarias, que representan las condiciones reales del terreno y permiten al modelo aprender patrones robustos y generalizables para la selección óptima de localizaciones.

5.3 Dataset y preprocesamiento

Las variables seleccionadas en la etapa anterior se organizan en un nuevo y único dataset que será el insumo para el entrenamiento de la red neuronal. El resultado es un archivo en formato CSV, el cual es importado a la plataforma de programación en línea Google Colaboratory. Esta plataforma ofrece una versión gratuita de Jupyter Notebook, un editor de código Python. Los datos deben ser preparados de manera conveniente para el posterior entrenamiento. Se llevan a cabo dos actividades principales sobre el dataset importado. Primero, se identifican y separan las variables independientes, que se asignan como entradas (inputs), y la variable dependiente, que se define como salida (output). Segundo, se realiza la normalización de los valores a un rango entre 0 y 1. Esta operación homogeneiza las escalas de las variables y elimina las distintas unidades de medida, lo que optimiza y acelera el proceso de entrenamiento del algoritmo.

5.4 División de datos

Debido a que el problema de predicción del potencial fotovoltaico está directamente relacionado con la ubicación geográfica, se descartó la división tradicional aleatoria en conjuntos de entrenamiento, validación y prueba. Esto podría ocasionar que el conjunto de prueba y el de validación contengan datos muy cercanos al conjunto de entrenamiento. La naturaleza geográfica del problema hace que exista una correlación espacial; es decir, los datos de lugares cercanos geográficamente tienden a compartir características similares (clima, relieve, irradiancia y otras propiedades físicas).

Para evitar esta dependencia espacial, se aplicó una estrategia de validación cruzada espacial basada en bloques, la cual divide el territorio en bloques espaciales geográficamente separados. Este método agrupa la latitud y longitud en grupos o folds, donde cada fold utiliza regiones geográficas completamente independientes entre sí. Esto garantiza que los datos utilizados para probar el modelo provengan de regiones geográficas distintas a las utilizadas para el entrenamiento.

La Tabla 4 muestra la división espacial en 5 folds para un dataset de 7006 registros utilizando un tamaño de bloque de 0.5° ($\approx$ 50 km), lo cual se considera adecuado para energía solar, y un tamaño de lote de 64. Se obtuvieron

cinco subconjuntos de entrenamiento y validación espacialmente disjuntos, cada uno manteniendo una distribución equilibrada del total de 7006 registros.

Table 4: División de datos mediante validación cruzada espacial (GroupKFold).

Fold	train_count	val_count	train_batches	val_batches
1	5604	1402	88	22
2	5605	1401	88	22
3	5605	1401	88	22
4	5605	1401	88	22
5	5605	1401	88	22

Una vez establecida la partición espacial del dataset, el siguiente paso consiste en diseñar y entrenar la red neuronal artificial que será utilizada para modelar la relación entre las variables geográficas, topográficas y climáticas y la potencia fotovoltaica generada.

5.5 Creación del modelo

En esta sección se describe la arquitectura propuesta de la red neuronal para la predicción de la potencia fotovoltaica. La arquitectura del modelo se basa en una estructura Multilayer Perceptron (MLP) diseñada para una tarea de regresión. La Figura 9 muestra de manera esquemática la configuración del MLP utilizado en este proyecto.

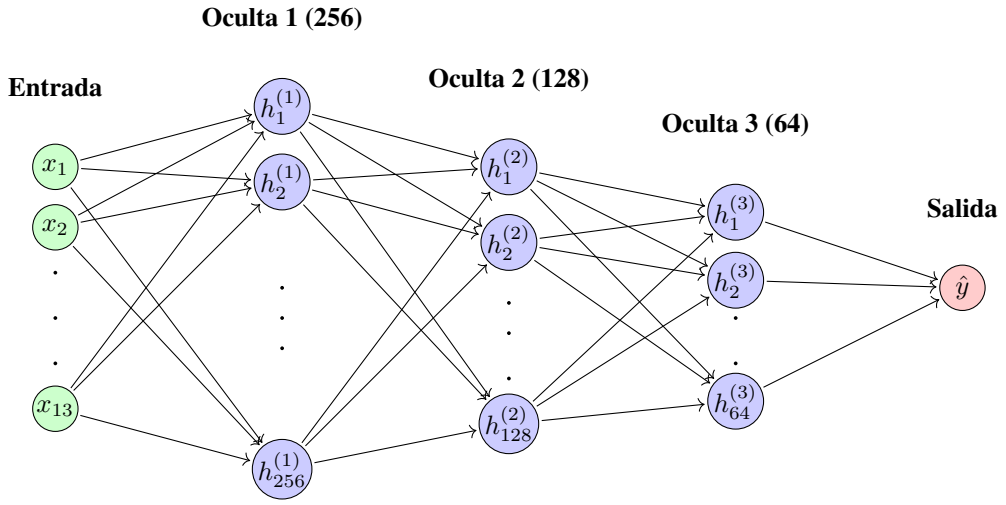


Figure 9: Arquitectura de la red neuronal artificial diseñada para la tarea de predicción de potencia PV.

Los datos ingresan a la red a través de una capa de entrada compuesta por 13 nodos, cada uno correspondiente a una de las variables independientes seleccionadas. Posteriormente, la información es procesada mediante tres capas ocultas densas con 256, 128 y 64 neuronas, respectivamente. En cada una de estas capas se emplea la función de activación ReLU, apropiada para capturar relaciones no lineales entre las variables. A continuación, se aplican las técnicas de Batch Normalization y Dropout, las cuales contribuyen a la estabilización del entrenamiento y a la reducción del sobreajuste al normalizar las activaciones y desactivar aleatoriamente un subconjunto de neuronas en cada iteración. Finalmente, una capa de salida con una única neurona, encargada de estimar la potencia fotovoltaica generada a partir de los valores de entrada, produciendo así una predicción continua coherente con la naturaleza del problema.

5.6 Entrenamiento

El entrenamiento es la fase central del ML. A partir del dataset previamente preparado, el algoritmo aprende a mapear las entradas y su respectiva salida. El proceso de entrenamiento tiene como objetivo que la red neuronal aprenda la relación funcional (mapeo) entre las variables geográficas, topográficas y climáticas, y la potencia fotovoltaica generada. Previamente, es necesario configurar ciertos valores denominados hiperparámetros, los cuales controlan este proceso. La Tabla reúne los hiperparámetros definidos para este estudio, que fueron cuidadosamente seleccionados con base en experimentación preliminar y en buenas prácticas para problemas de regresión con datos ambientales.

Table 5: Hiperparámetros de entrenamiento de la red neuronal.

Hiperparámetro	**Valor**
Dimensión de la entrada	13
Dimensión de la salida	1
Tamaño de lote	64
Función de activación	ReLU
Función de pérdida (Loss)	Mean Absolute Error (MAE)
Optimizador	Adam
Tasa de aprendizaje (lr)	1×10^{-3}
Métrica de evaluación	MAE
Dropout	0.2
EarlyStopping	val_loss
Folds	5
Épocas	300

La red neuronal recibe una entrada de dimensión 13, correspondiente al número total de variables independientes seleccionadas. Estos valores son procesados a través de tres capas ocultas, cuyas neuronas combinan la información de entrada y aplican la función de activación ReLU para introducir no linealidad y permitir que el modelo aprenda relaciones complejas entre las variables. Para evaluar el desempeño del modelo, se emplea el error absoluto medio (MAE), una métrica robusta que cuantifica la diferencia promedio entre las predicciones y los valores reales de potencia fotovoltaica.

Con el fin de minimizar esta función de error, se utiliza el optimizador Adam, un algoritmo derivado del descenso del gradiente que ajusta los parámetros del modelo de manera iterativa mediante una tasa de aprendizaje adaptable. Durante el proceso de entrenamiento, los pesos y sesgos se actualizan automáticamente mediante el algoritmo de retropropagación (backpropagation), permitiendo que la red neuronal mejore progresivamente su capacidad predictiva. Además, se incorporó la técnica de early stopping, que detiene el entrenamiento cuando la mejora en la función de pérdida deja de ser significativa, evitando así un sobreajuste a los datos de entrenamiento.

El modelo se entrena utilizando el esquema de validación cruzada espacial, lo que permite evaluar su capacidad real de generalización en regiones geográficas no observadas. Durante cada fold, la red neuronal aprende a partir de los datos de entrenamiento y es evaluada en el conjunto de validación espacialmente independiente. Este procedimiento asegura una estimación robusta del error y permite seleccionar la mejor configuración del modelo antes de entrenar una versión final utilizando el dataset completo.

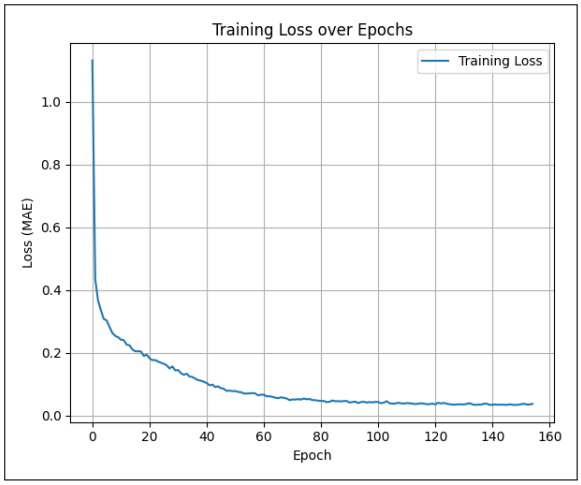
5.7 Evaluación

La evaluación del modelo se realiza en dos etapas complementarias. En primer lugar, se aplica una validación cruzada espacial de 5 folds, técnica adecuada cuando los datos están geográficamente distribuidos y se busca evitar el sobreajuste por proximidad espacial. En cada pliegue, el modelo se entrena en un conjunto de regiones y se evalúa en regiones distintas, obteniendo un valor de MAE por fold. El resultado final de esta etapa se resume mediante el MAE promedio y su desviación estándar, lo que permite medir la capacidad del modelo para generalizar a nuevas ubicaciones geográficas no vistas. La Tabla 6 resume la métrica de evaluación MAE.

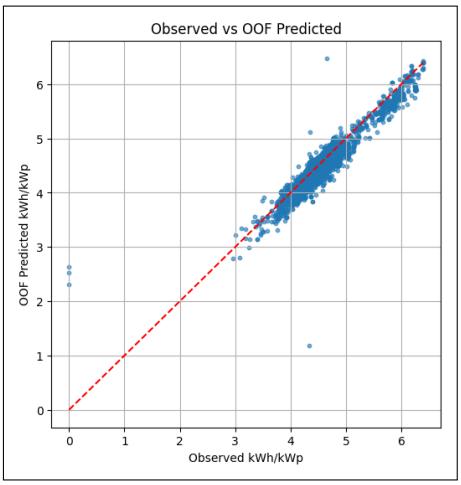
Table 6: Valores de la métrica de evaluación MAE durante el entrenamiento.

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	MAE mean +/- std
0.109	0.090	0.107	0.092	0.100	0.100 +/- 0.008

En segundo lugar, una vez comprobado el desempeño del modelo mediante la validación cruzada, se entrena un modelo final utilizando el conjunto completo de datos disponibles. Este modelo aprovecha toda la información para mejorar la capacidad predictiva y será el utilizado para generar las predicciones en la grilla espacial de interés. El proceso de entrenamiento se realiza con la misma configuración de hiperparámetros, pero sin conjunto de validación, ya que la calidad del modelo ya fue demostrada durante la validación cruzada. La Figura 10 muestra las gráficas de rendimiento del modelo final.



(a) Curva de aprendizaje.



(b) Comparación de lo observado y lo real.

Figure 10: Curva de aprendizaje del modelo y comparación de los valores observados y reales.

5.8 Predicción

A partir del modelo entrenado se puede estimar la potencial producción fotovoltaica de una planta solar en un determinado lugar sin necesidad de mediciones locales. Se puede usar esta información para generar un mapa de susceptibilidad, el cual responda: ¿Qué tan bueno es cada punto del territorio para instalar una planta solar? De esta manera, es posible identificar zonas favorables. Estas estimaciones deberían usarse como primera aproximación o preselección, y complementarse con estudios de campo antes de decisiones definitivas.

El proceso de predicción requiere la generación de una cuadrícula para un área geográfica definida por el usuario y, a continuación, ejecutar el modelo de predicción del potencial fotovoltaico en esta cuadrícula generada. Los pasos específicos son:

1. Definir parámetros del área geográfica: el usuario proporciona la latitud mínima/máxima, la longitud mínima/máxima y un tamaño de paso para la generación de la cuadrícula.
2. Generar las celdas de la cuadrícula (puntos de datos geoespaciales) según los parámetros definidos por el usuario.
3. Para cada punto de la cuadrícula, se rellenan la latitud y longitud y las demás columnas de las variables necesarias que utiliza el modelo.
4. Ejecutar el modelo de predicción del potencial fotovoltaico en la cuadrícula creada. Esto implica escalar sus variables utilizando el escalador preentrenado, predecir el potencial fotovoltaico para cada punto y guardar los resultados en un archivo GeoJSON.
5. Generar un mapa interactivo con superposición de satélite, el cual se guarda en formato HTML para ser visualizado en un navegador.

La Figura 11 muestra la predicción realizada por el modelo para el área geográfica definida con latitud mínima de $-35,0$, latitud máxima de $-30,0$, longitud mínima de $-70,0$, longitud máxima de $-65,0$ y un tamaño de paso de $0,1$. Se generó correctamente una cuadrícula con 2652 puntos para esta zona y se aplicó el modelo de predicción del potencial fotovoltaico a estos puntos y las predicciones resultantes se despliegan en el mapa satelital.

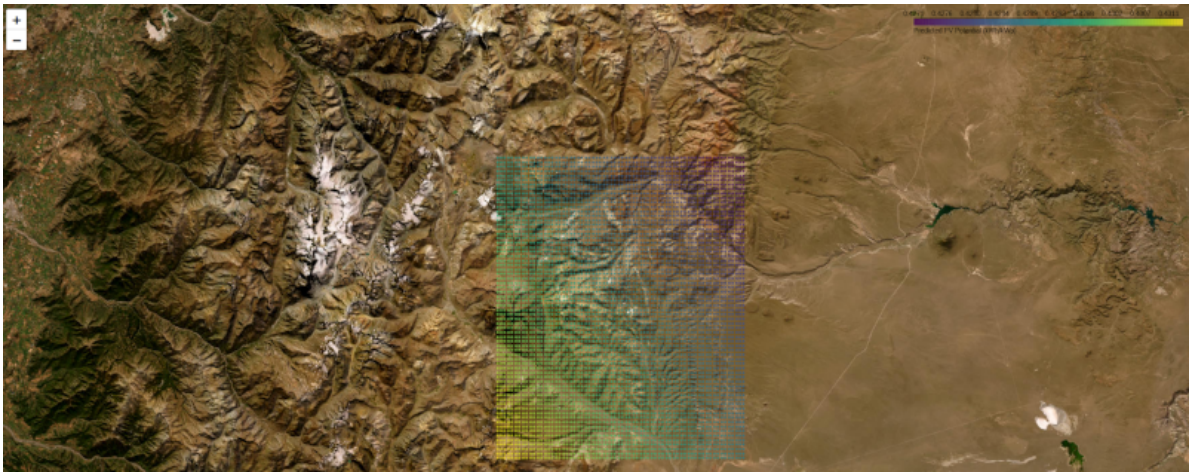


Figure 11: Predicción de potencial PV para una zona geográfica.

6 Conclusiones

El método desarrollado para la creación del dataset de plantas solares constituye un proceso integral, sistemático y completamente replicable que garantiza la obtención de información confiable, homogénea y científicamente válida. La claridad del flujo de trabajo y la precisión de sus etapas permiten que este proceso pueda ser reproducido, ampliado o adaptado por futuros investigadores en el campo de las energías renovables. La búsqueda bibliográfica exhaustiva que se realiza al inicio fundamenta la selección de atributos de manera objetiva y basada en evidencia, evitando la inclusión de parámetros irrelevantes. Posteriormente, la definición de los atributos y de la estructura del dataset, permite organizar de manera clara y coherente las variables en categorías fundamentales, con lo cual se asegura trazabilidad, reproducibilidad y compatibilidad con análisis estadísticos y modelos de aprendizaje automático. El resultado es un método es vasto porque integra análisis documental, herramientas SIG, procesamiento de datos climáticos y validación técnica, garantizando que el dataset no solo sea completo y confiable, sino también adecuado para aplicaciones avanzadas como la selección óptima de emplazamientos mediante deep learning.

El dataset construido representa un recurso de alto valor científico y técnico, resultado de un proceso riguroso de recopilación, depuración, estandarización, organización, y verificación de información proveniente de múltiples fuentes confiables. Su estructura integra de manera coherente variables de identificación, ubicación, topografía, clima y operación, permitiendo una caracterización completa y multidimensional de las plantas solares fotovoltaicas existentes. Este data set llena un vacío existente en el ámbito de la investigación científica, constituyéndose en una base de datos robusta, precisa y altamente funcional. La incorporación de atributos como pendiente, curvatura, aspecto, distancia a vías y aptitud solar, en conjunto con datos climáticos como irradiancia, temperatura, humedad y viento, convierten al dataset en una herramienta integral para estudios de localización óptima, modelado energético, simulaciones geoespaciales, aplicaciones de inteligencia artificial, entre muchas otras. Esta combinación de variables no suele encontrarse reunida en una sola fuente, lo cual marca una diferencia sustancial respecto de bases de datos fragmentadas y parciales disponibles en el ámbito público.

El modelo de inteligencia artificial utilizado demostró ser adecuado para capturar las relaciones no lineales entre variables geográficas, topográficas, climáticas y de potencia generada por plantas solares. Las técnicas como Batch Normalization, Dropout, ReLU, el optimizador Adam y Early Stopping, permitieron un entrenamiento eficiente sin sobreajuste, lo que a su vez evidencia que el dataset cuenta con información relevante y bien estructurada para alimentar modelos predictivos de energía solar. La implementación de validación cruzada espacial GroupKFold fue un elemento clave, ya que permitió evaluar el modelo en zonas no observadas, reduciendo el sesgo espacial y garantizando una capacidad de generalización realista. Todo esto en conjunto confirma que el modelo es útil no solo para predecir potencia, sino también como base metodológica para aplicaciones de localización óptima de plantas solares.

7 Disponibilidad y Acceso al Dataset

Los datos y el código que respaldan los hallazgos del presente estudio están disponibles en los siguientes enlaces:
Modelo de predicción: https://github.com/cimejia/solarPV/blob/main/ML-model/solarpvprediction.ipynb
Conjunto de datos: https://github.com/cimejia/solarPV/blob/main/Dataset/Dataset_Mundial.xlsx

References

[1] L. W. Lijiang. A systematic review of photovoltaic power plant site selection approaches: Literature retrieval and analysis method based on chatgpt and deepseek. *Energy Reports*, 14:1993–2014, 2025.

[2] S. S. Mekhilef. A review on solar energy use in industries. *Renewable and Sustainable Energy Reviews*, 15(4):1777–1790, 2011.

[3] Juraj Betak, Marek Caltik, Tomas Cebecauer, Daniel Chrkavy, Branislav Erdelyi, Konstantin Rosina, Marcel Suri, and Nada Suriova. Global photovoltaic power potential by country, 2020.

[4] H. L. Fang. Sustainable site selection for photovoltaic power plant: An integrated approach based on prospect theory. *Energy Conversion and Management*, 174:755–768, 2018.

[5] T. Sharma and A. Sarin. Multi-criteria decision making for solar site selection in punjab, india: An evaluation of site suitability using hybrid mcdm techniques towards the goal of sustainable energy development. *Results in Engineering*, 27:106288, 2025.

[6] G. L. Goncalves, L. M. de A. L. Gaudencio, R. de Oliveira, and R. S. do Nascimento. Sustainable planning of solar plants: bibliometry and integration of gis-multicriteria methods. *Sustainable Futures*, 10:100993, 2025.

[7] N. L. Rane. Gis-based multi-influencing factor (mif) application for optimal site selection of solar photovoltaic power plant in nashik, india. *Environmental Sciences Europe*, 36(1):5, 2024.

[8] Pablo Benalcazar, Aleksandra Komorowska, and Jacek Kamiński. A gis-based method for assessing the economics of utility-scale photovoltaic systems. *Applied Energy*, 353, 1 2024.

[9] Jeffrey T Dellosa and Eleonor V Palconit. www.etasr.com dellosa & palconit: Resource assessment of a floating solar photovoltaic (fspv) system with .... Technical report, 2022.

[10] İ. Karadöl and R. Yıldırım. Location choice in solar power plants by applying meteorological data to multi-criteria decision-making method. *Engineering Applications of Artificial Intelligence*, 152:110766, 2025.

[11] Chukwuebuka Joseph Ejiyi, Dongsheng Cai, Dara Thomas, Sandra Obiora, Emmanuel Osei-Mensah, Caroline Acen, Francis O. Eze, Francis Sam, Qingxian Zhang, and Olusola O. Bamisile. Comprehensive review of artificial intelligence applications in renewable energy systems: current implementations and emerging trends. *Journal of Big Data*, 12, 12 2025.

[12] Y. Sun, Y. Li, R. Wang, and R. Ma. Dynamic geospatial modeling of solar energy expansion potential in china: Implications for national-level optimization of solar photovoltaic plant layouts. *Energy*, 333:137433, 2025.

[13] Héctor Felipe Mateo Romero, Miguel Ángel González Rebollo, Valentín Cardeñoso-Payo, Victor Alonso Gómez, Alberto Redondo Plaza, Ranganai Tawanda Moyo, and Luis Hernández-Callejo. Applications of artificial intelligence to photovoltaic systems: a review. *Applied Sciences*, 12(19):10056, 2022.

[14] Anqi Li, Luling Liu, Shijie Li, Xihong Cui, Xuehong Chen, and Xin Cao. Global photovoltaic solar panel dataset from 2019 to 2022. *Scientific Data 2025 12:1*, 12:637–, 4 2025.

[15] Byron Guerrero-Rodriguez, Jose Garcia-Rodriguez, Jaime Salvador, Christian Mejia-Escobar, Shirley Cadena, Jairo Cepeda, Manuel Benavent-Lledo, and David Mulero-Perez. Improving landslide prediction by computer vision and deep learning. *Integrated Computer-Aided Engineering*, 31(1):77–94, 2024.

[16] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. The world-wide web. *Commun. ACM*, 37(8):76–82, aug 1994.

[17] James M Gillies, James Gillies, R Cailliau, et al. *How the Web was born: The story of the World Wide Web*. Oxford University Press, USA, 2000.